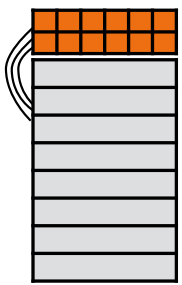
Data center architecture

Ankit Singla

ETH zürich

A server rack

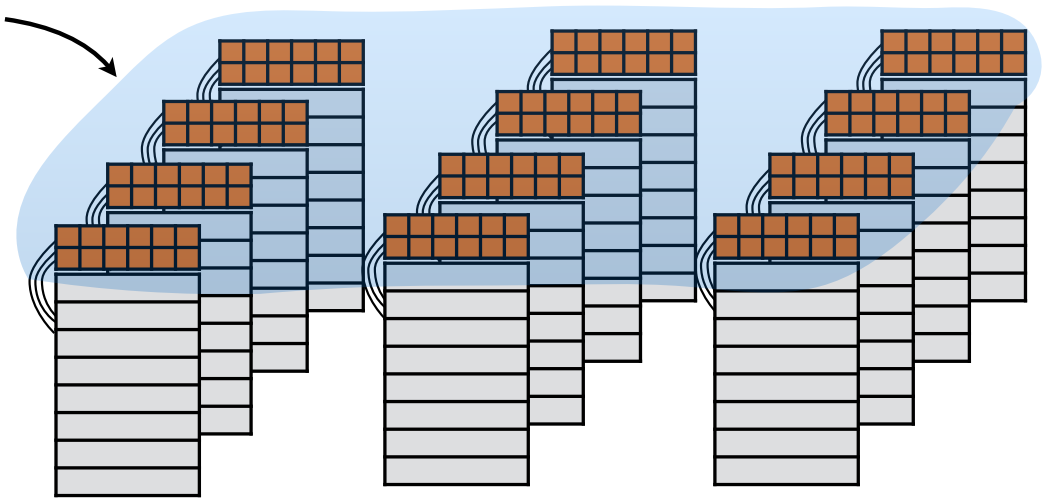


A top-of-rack switch

A rack of servers

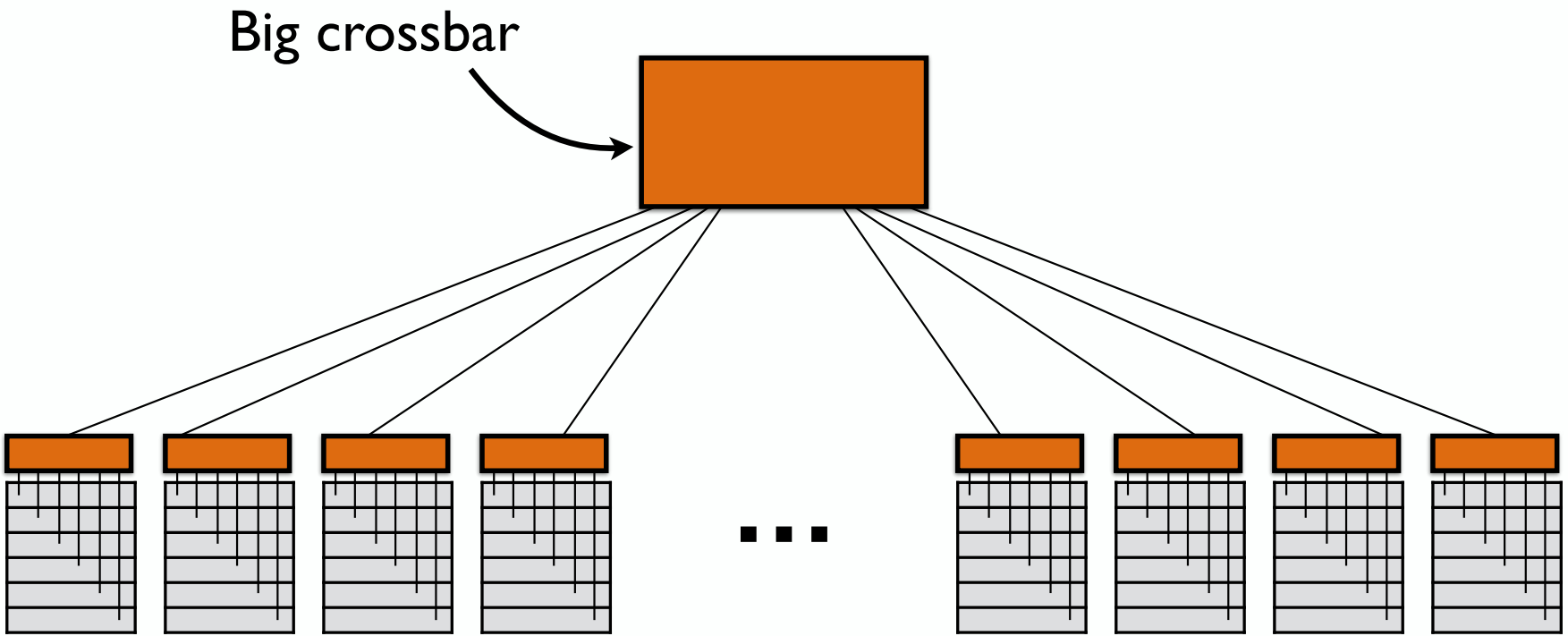
Lots of racks

How to network the racks?

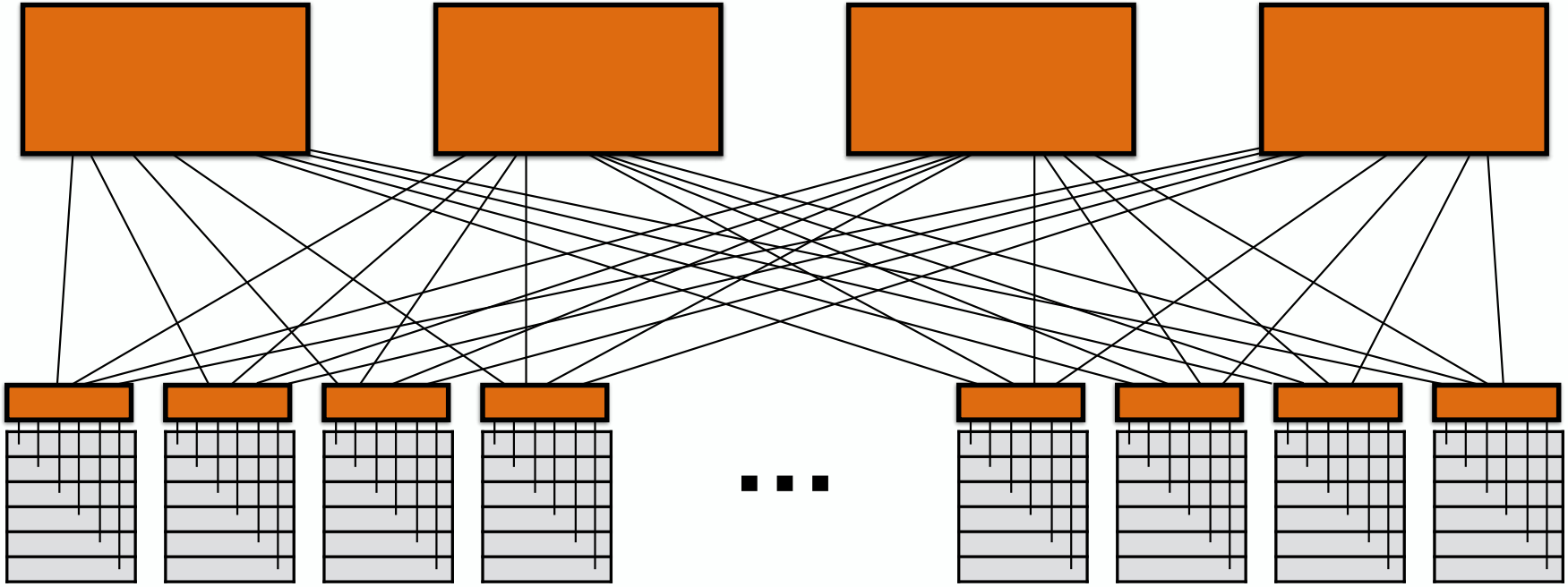


Facebook: machine-machine traffic "doubling at an interval of less than a year"

“Big switch” approach

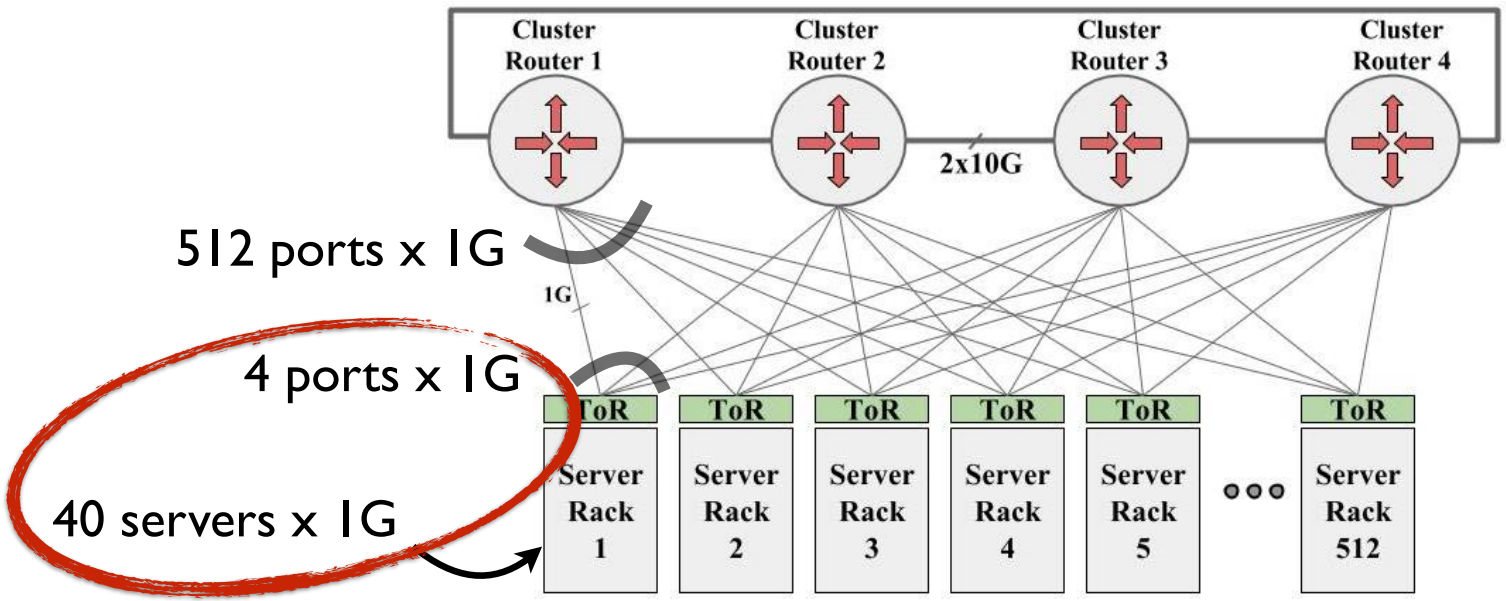


“Big switch” approach



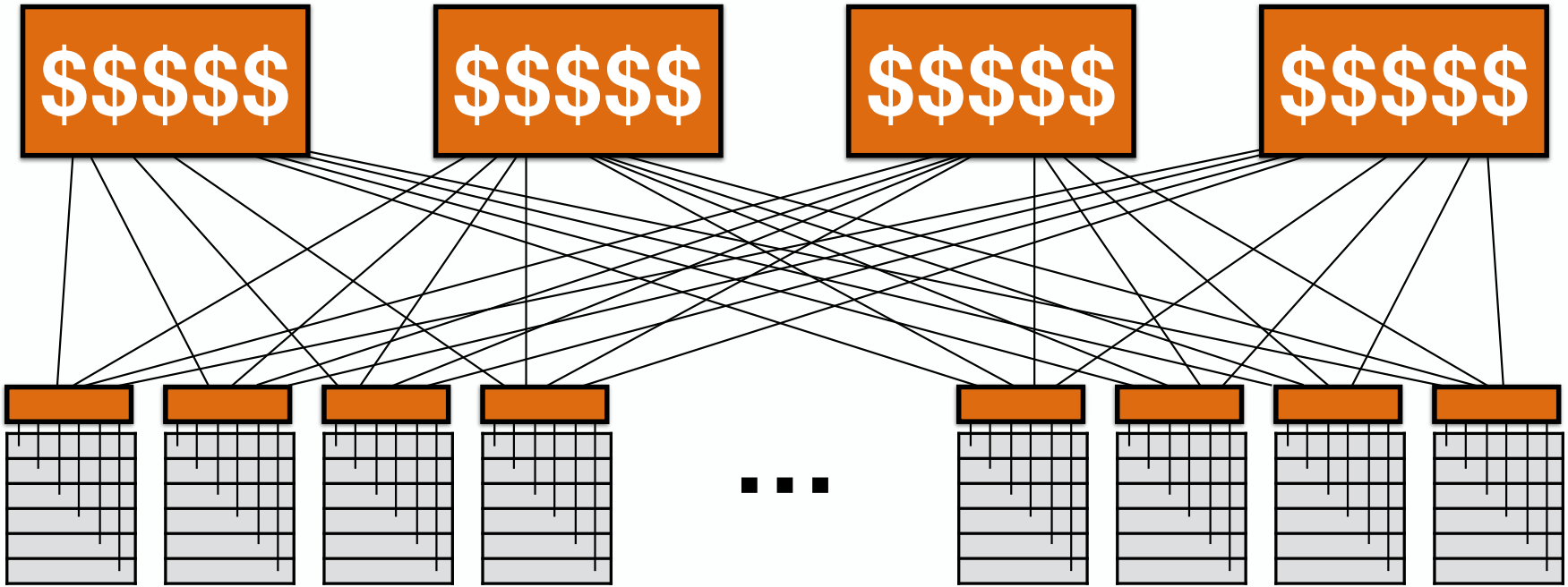
Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network

Arjun Singh, Joon Ong, Amit Agarwal, Glen Anderson, Ashby Armistead, Roy Bannon, Seb Boving, Gaurav Desai, Bob Felderman, Paulie Germano, Anand Kanagala, Jeff Provost, Jason Simmons, Eiichi Tanda, Jim Wanderer, Urs Hölzle, Stephen Stuart, and Amin Vahdat
Google, Inc.

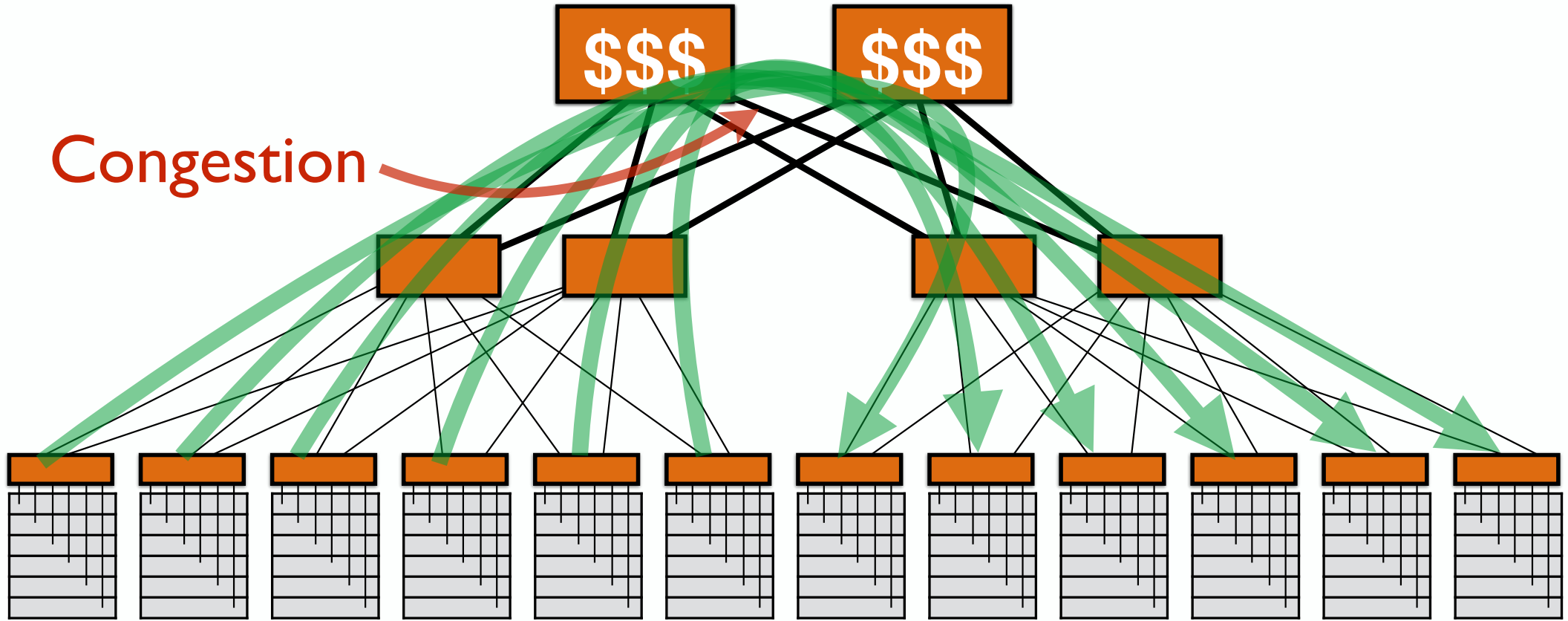


10:1 "oversubscription"

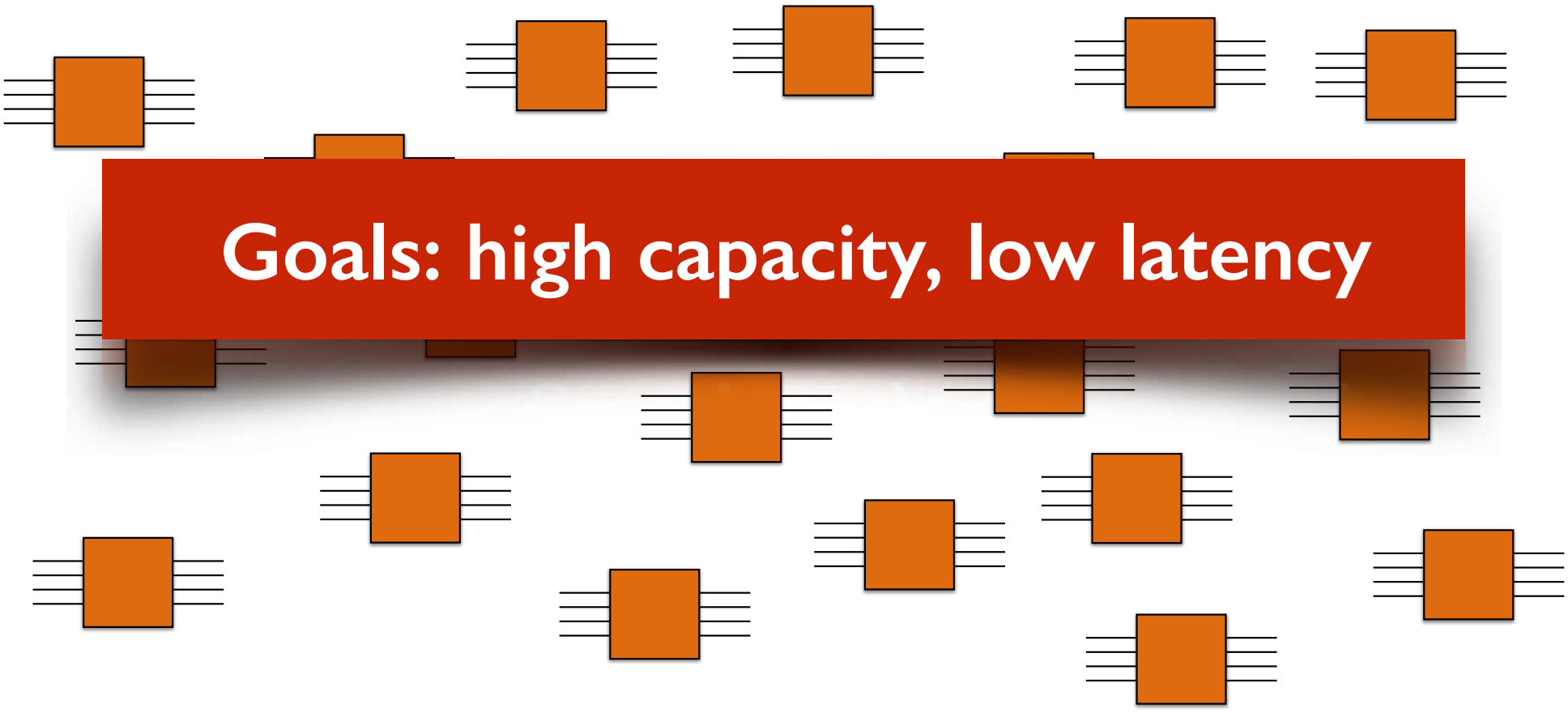
“Big switch” approach



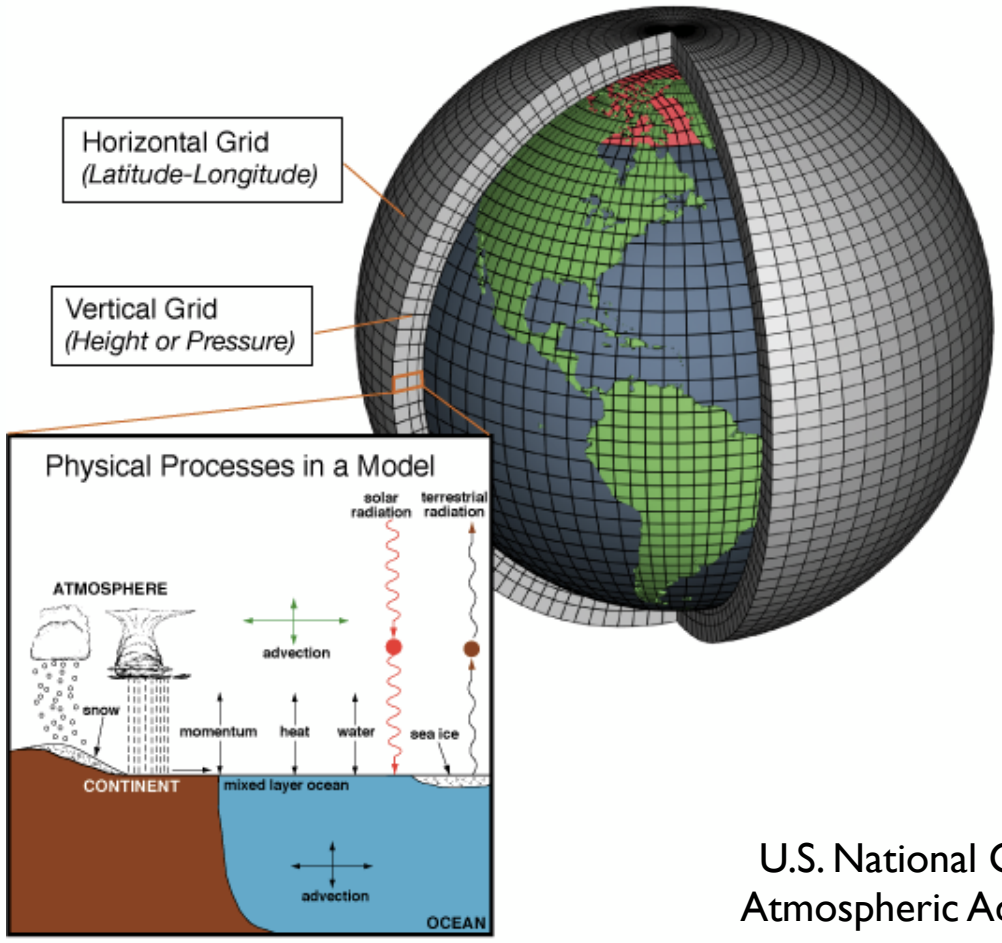
Alternative: tree network



Connect many cheap, identical switches?

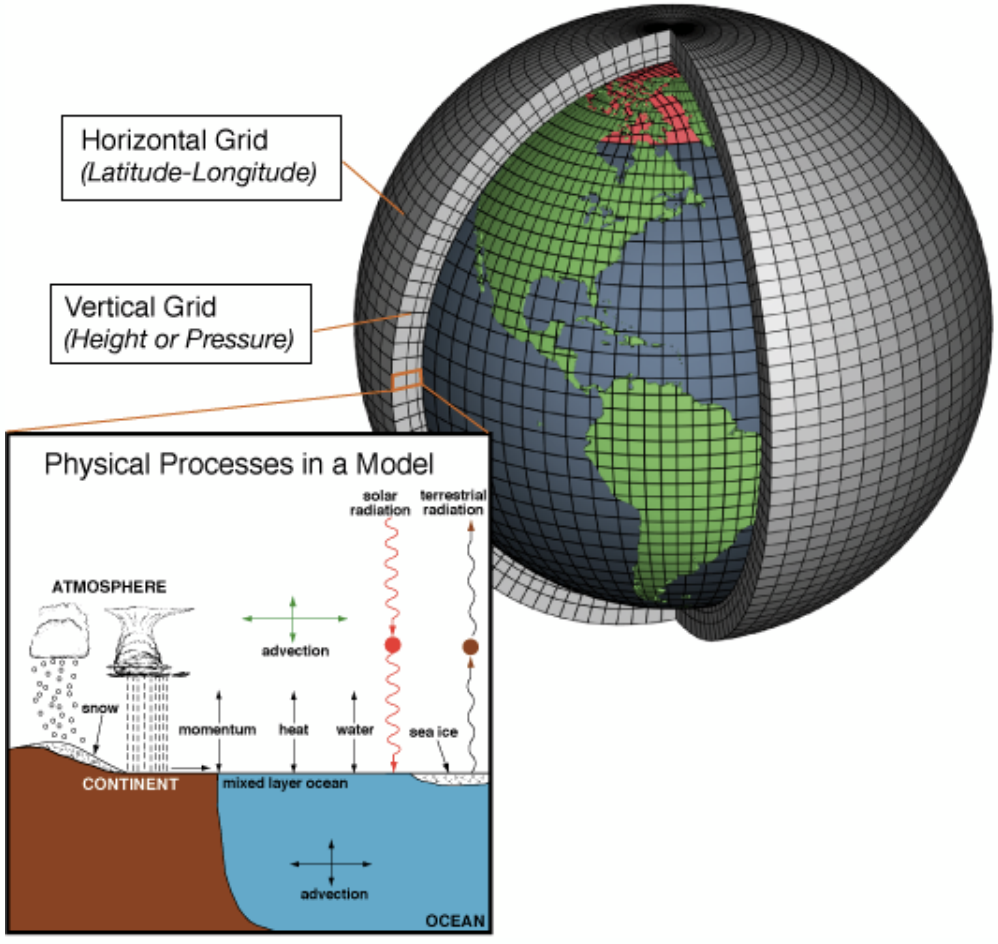


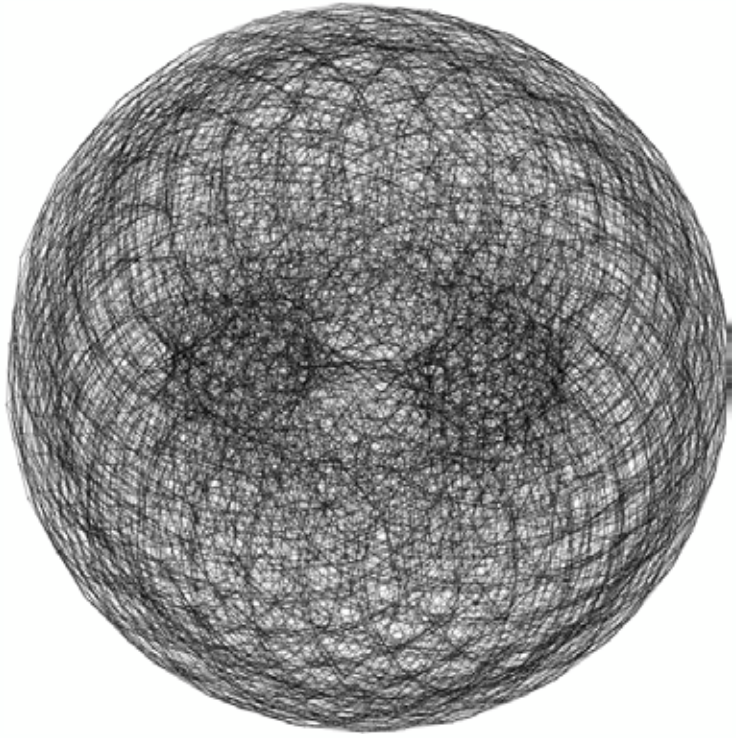
If you know your application …



U.S. National Oceanic and Atmospheric Administration

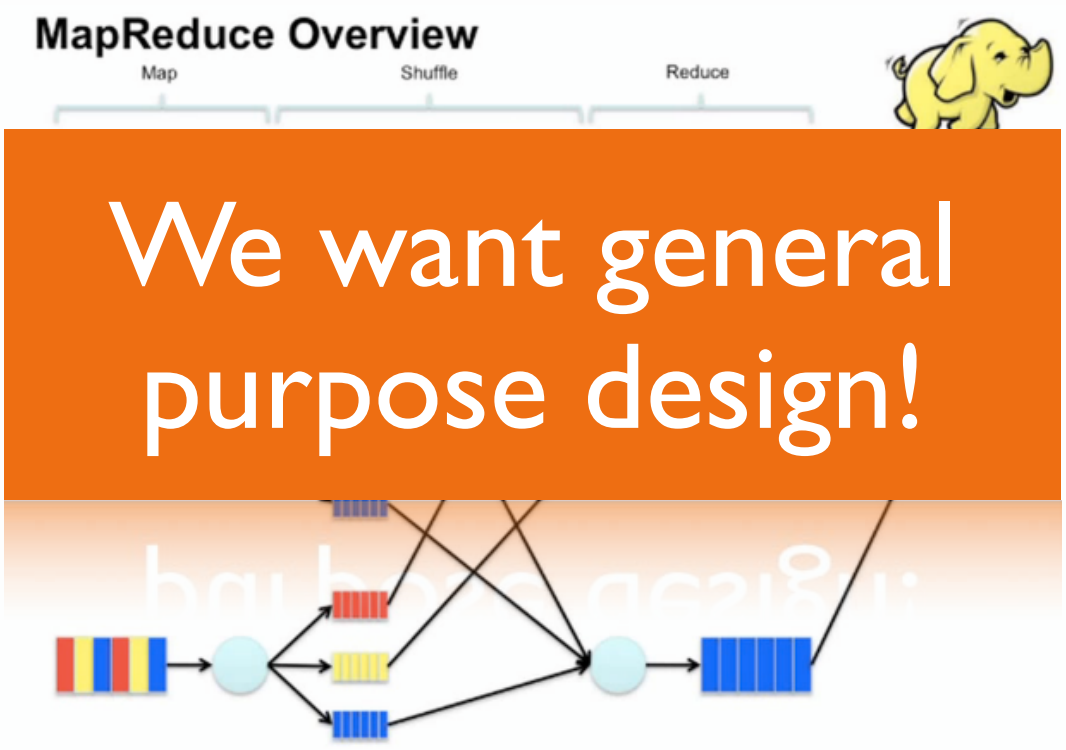
… design for it



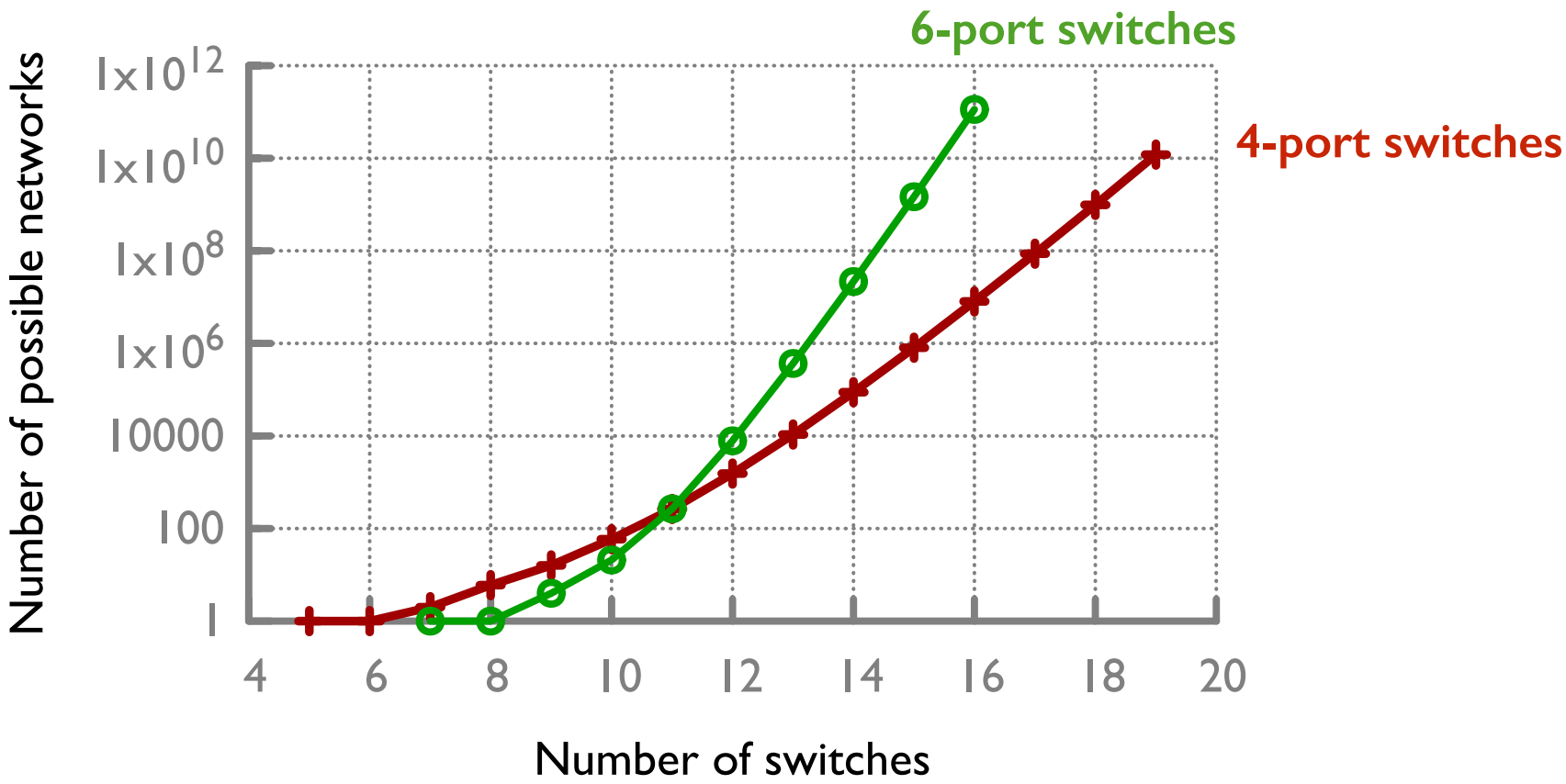


"Hopper", NERSC

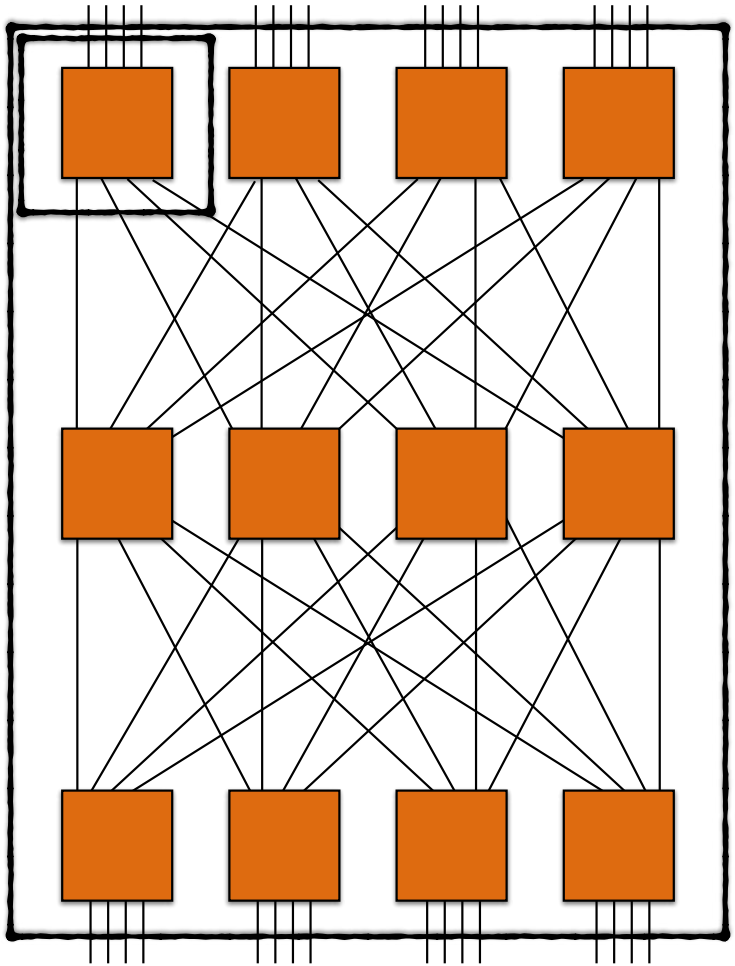
But, other apps may not work well …



What's so hard about this?



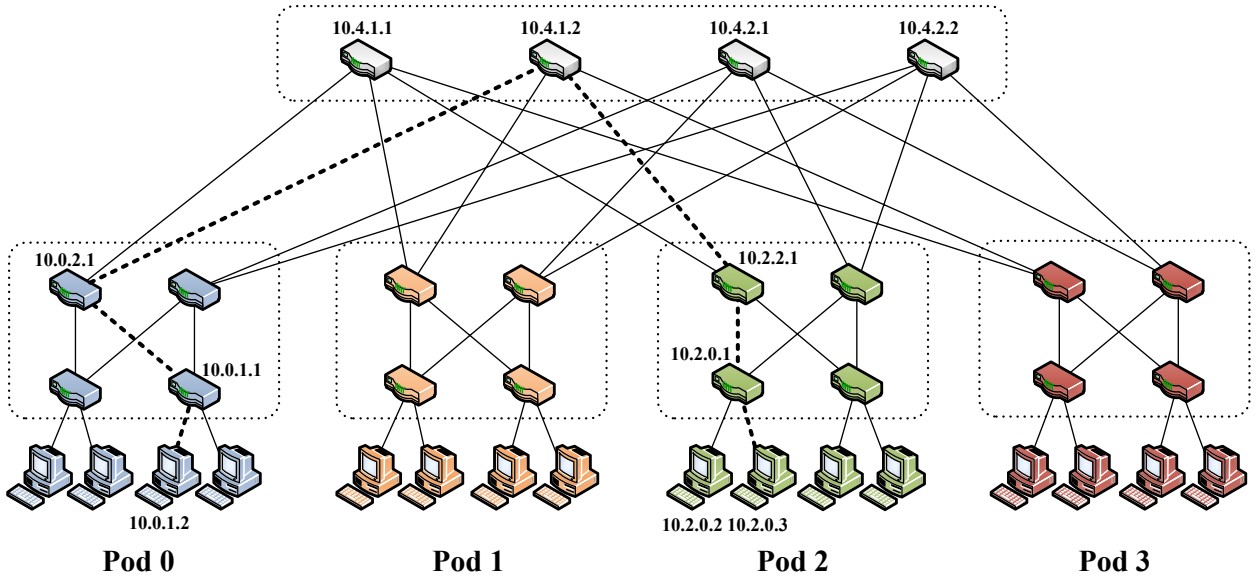
Clos networks



Use small, cheap elements to build large networks!!

Fat-tree

Fat-tree network

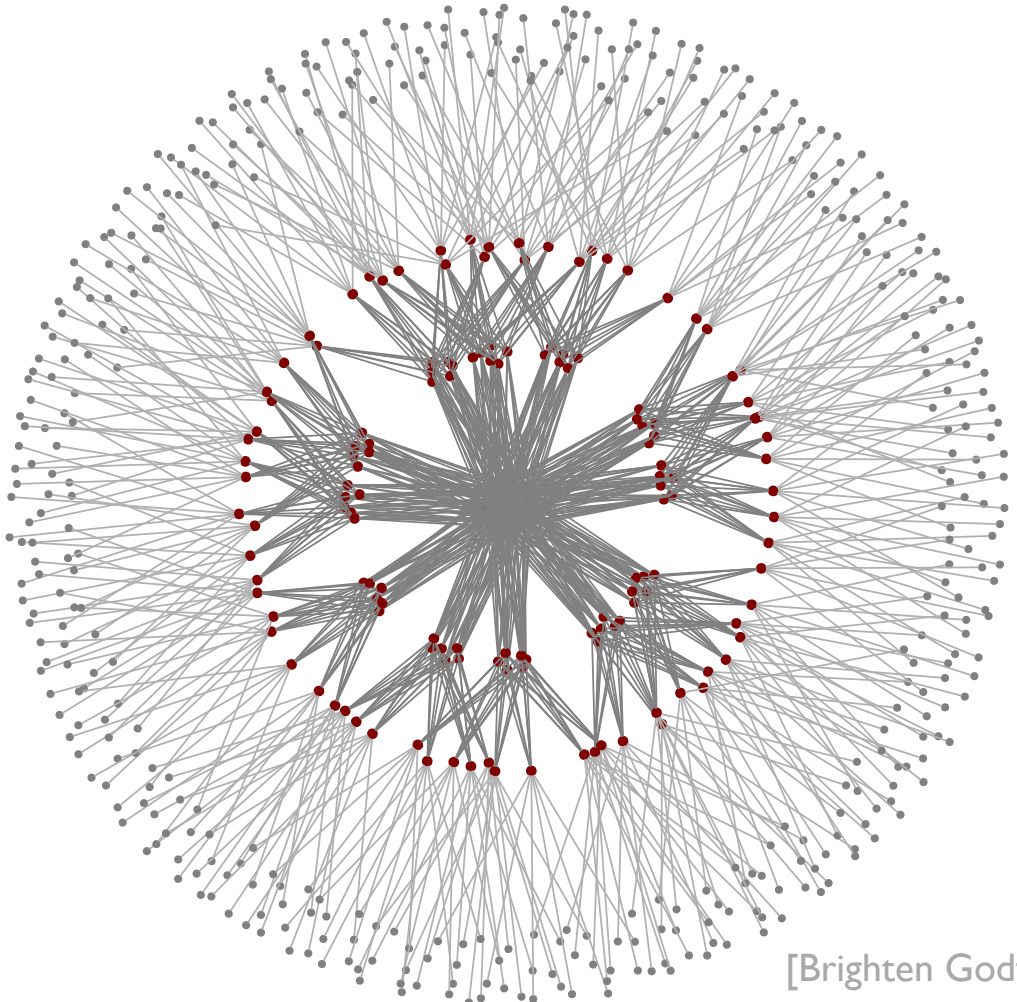


ACM SIGCOMM, 2008

A Scalable, Commodity Data Center Network Architecture

Mohammad Al-Fares Alexander Loukissas Amin Vahdat

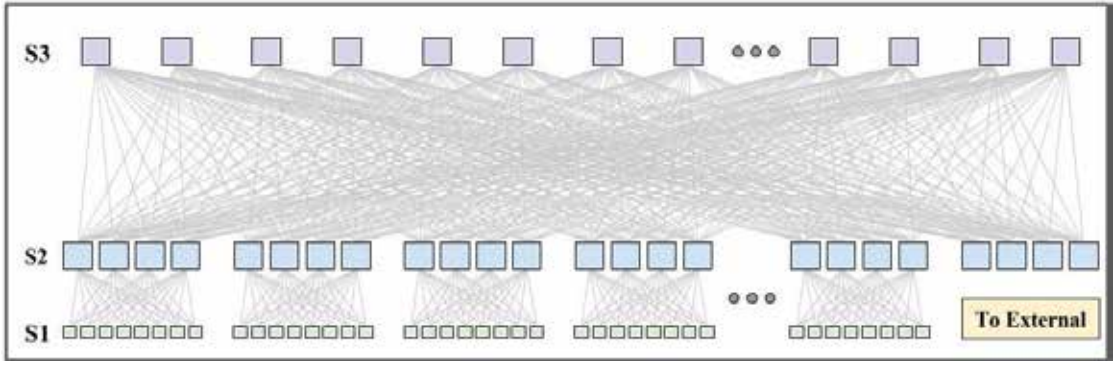
Fat-tree network



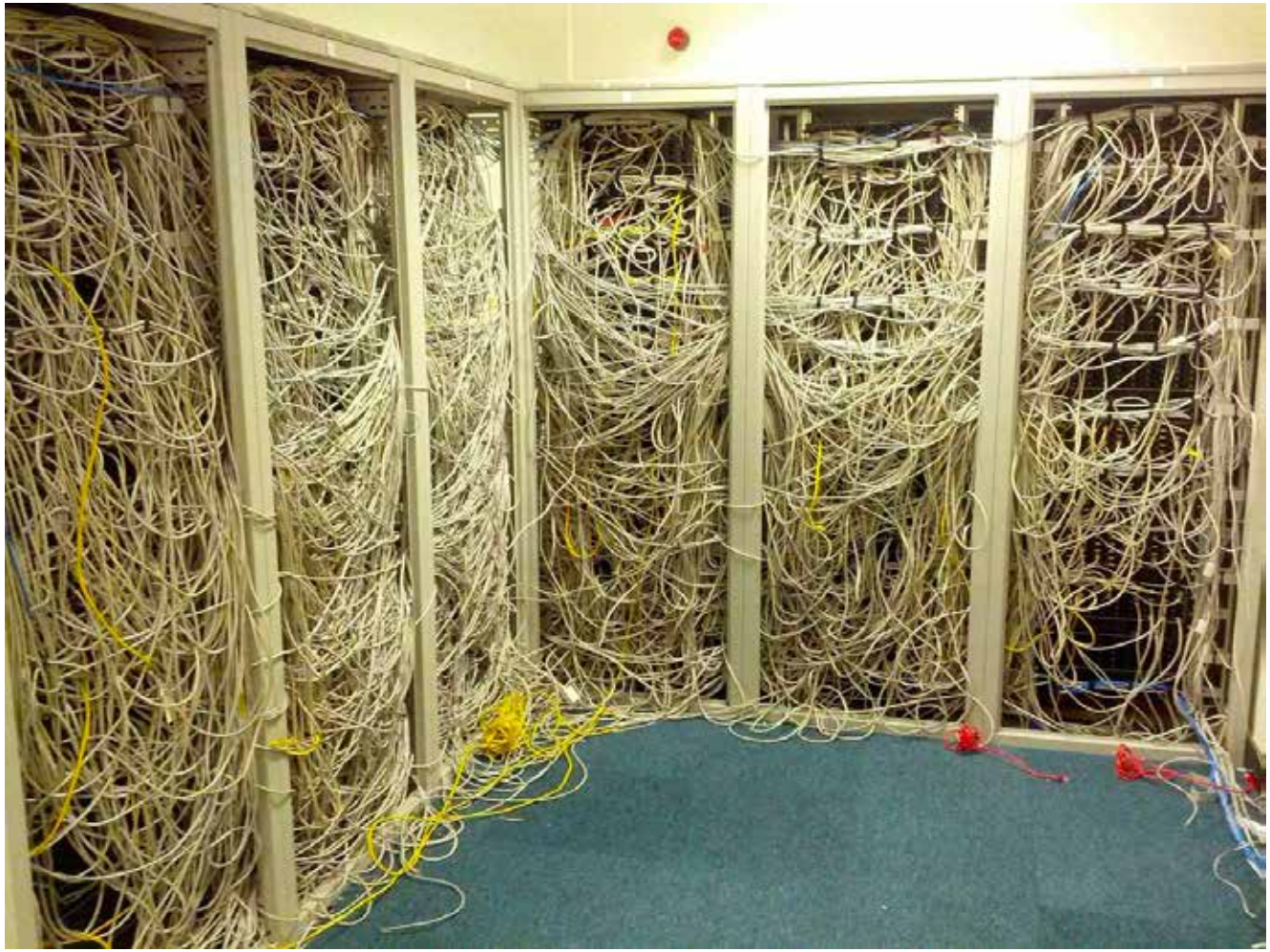
[Brighten Godfrey]

Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network

Arjun Singh, Joon Ong, Amit Agarwal, Glen Anderson, Ashby Armistead, Roy Bannon, Seb Boving, Gaurav Desai, Bob Felderman, Paulie Germano, Anand Kanagala, Jeff Provost, Jason Simmons, Eiichi Tanda, Jim Wanderer, Urs Hölzle, Stephen Stuart, and Amin Vahdat
Google, Inc.



[David Samuel Robbins, gettyimages.ch]



[@AlexCWheeler, Twitter]

Explore Wiring Fail, Wiring Jobs, and more!

Computers | Cable | Thoughts | The spider | Wells | Lord of the rings | Need to | Yellow

Wiring Fail | Wiring Jobs | Safe Wiring

Poor data center cable management. I'm expecting Shelob the spider from Lord of the Rings to emerge any moment.

See More

9 1

Datacenter Di | Datacenter Google

Un giro nei #datacenter di #Google

See More

12

Horrible Data | Wiring Disasters

Aaaah! What a horrible data center disaster.

See More

by Eric Brandwine

1

Messy Cable Closets & Serv

13 Pins 43 Followers

Cable | To work | Search | Wire | Tech | Try again | Gazebo | Google

Room Disasters | Computer Disasters

Server Room disaster

See More

3

Room Nightmares | Cabling Nightmares

Real-world server room nightmares

See More

2

Messy Cable | Worst Se

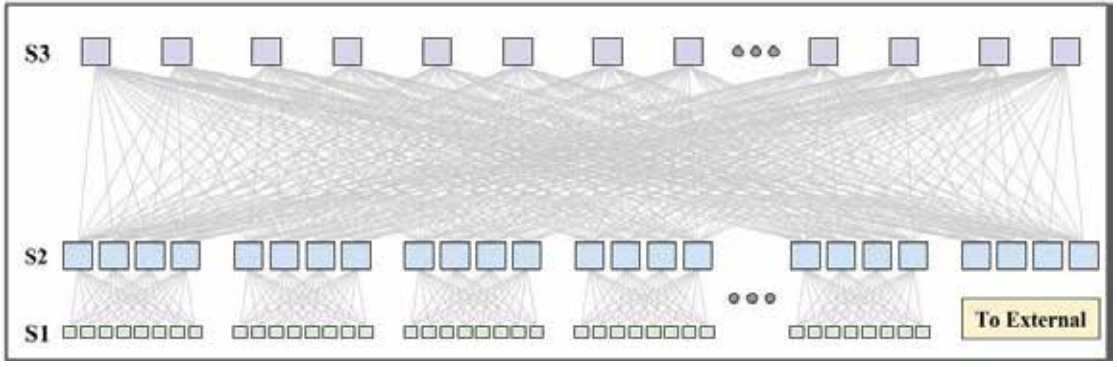
Real-world server roo

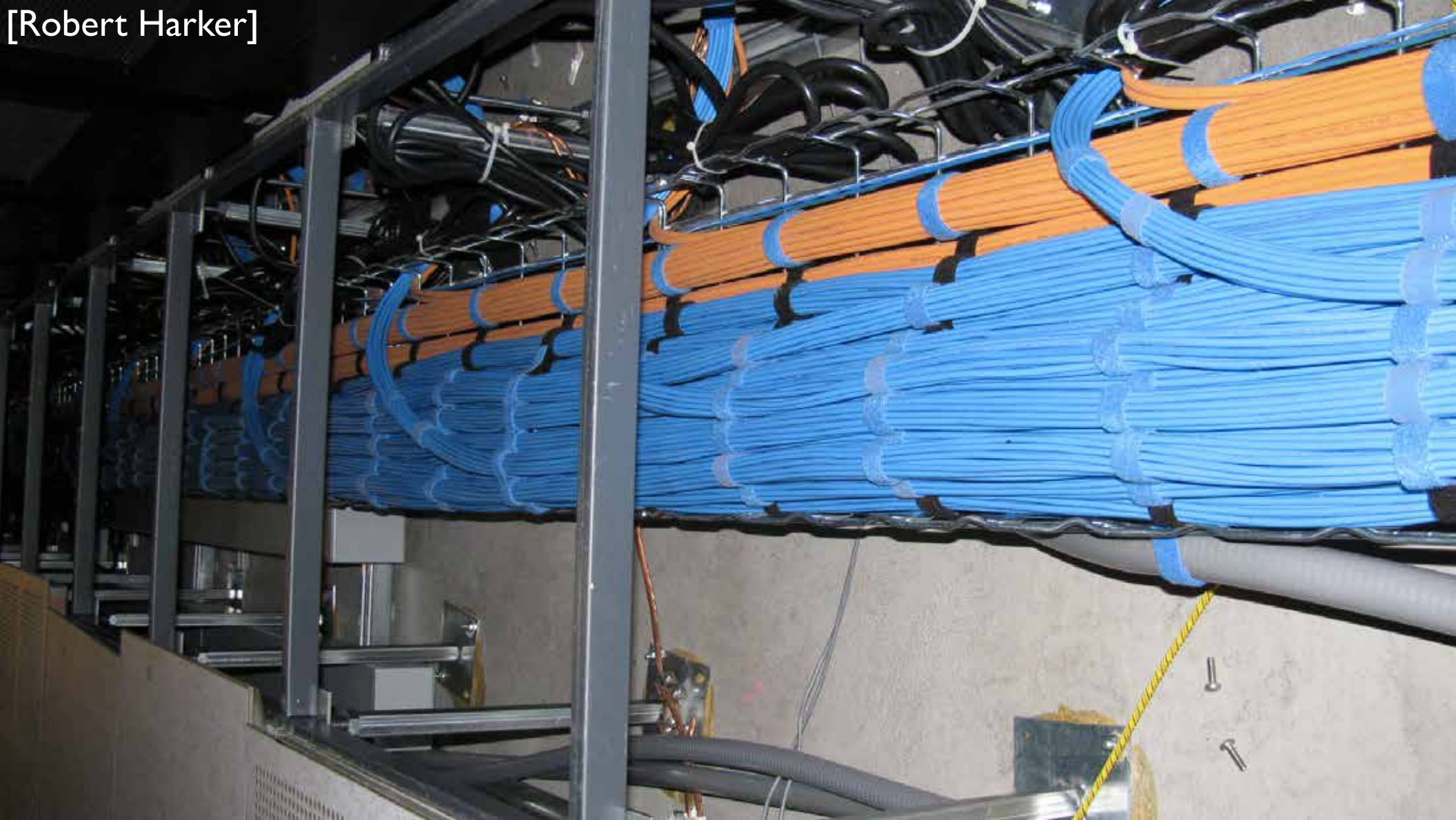
See More

1

Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network

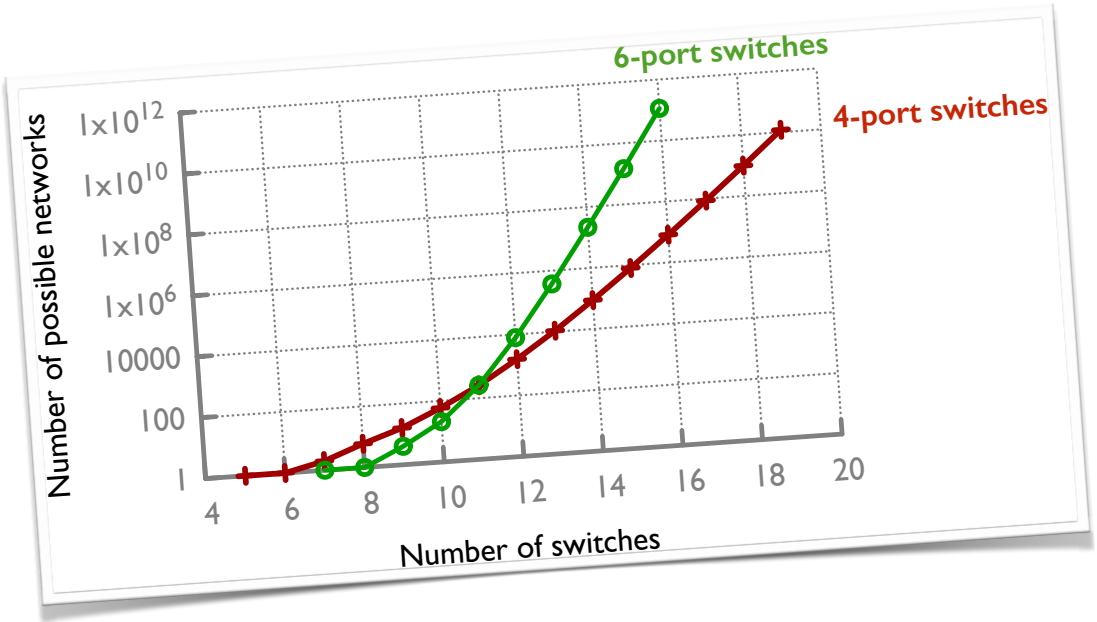
Arjun Singh, Joon Ong, Amit Agarwal, Glen Anderson, Ashby Armistead, Roy Bannon, Seb Boving, Gaurav Desai, Bob Felderman, Paulie Germano, Anand Kanagala, Jeff Provost, Jason Simmons, Eiichi Tanda, Jim Wanderer, Urs Hölzle, Stephen Stuart, and Amin Vahdat
Google, Inc.




[Robert Harker]

DC architecture

Other topologies besides fat-trees?



Surprise: picking randomly from this space works better than fat-trees!

Jellyfish: Networking Data Centers Randomly, NSDI 2012, Ankit Singla, Chi-Yao Hong, Lucian Popa, P. Brighten Godfrey

Deterministic expander constructions can also achieve these gains

Xpander: Towards Optimal-Performance Datacenters, CoNEXT 2016, Asaf Valadarsky, Gal Shahaf, Michael Dinitz and Michael Schapira

But do we need full bandwidth everywhere at all times?

25% is the high mark for overall utilization (?)

"number of highly utilized core links varies over time, but never exceeds 25%"

Network Traffic Characteristics of Data Centers in the Wild. Benson et al. ACM IMC 2010.

"the network traffic is very low; very few nodes exceed 0.0064MB/s" — Microsoft

ECHO: Recreating Network Traffic Maps for Datacenters with Tens of Thousands of Servers. Delimitrou et al. IEEE IISWC 2012.

"busiest 5% of the links seeing 23–46% utilization." — Facebook

Inside the Social Network's (Datacenter) Network. Roy et al. ACM SIGCOMM 2015.

"experienced high congestion drops as utilization approached 25%" — Google

Jupiter Rising: A Decade of Clos Topologies and Centralized Control in Google's Datacenter Network. Singh et al. ACM SIGCOMM 2015.

"46-99% of the rack pairs exchange no traffic at all" — Microsoft

ProjecToR: Agile Reconfigurable Data Center Interconnect. Ghobadi et al. ACM SIGCOMM 2016.

25% utilization of oversubscribed network may imply ~10% from a server perspective!

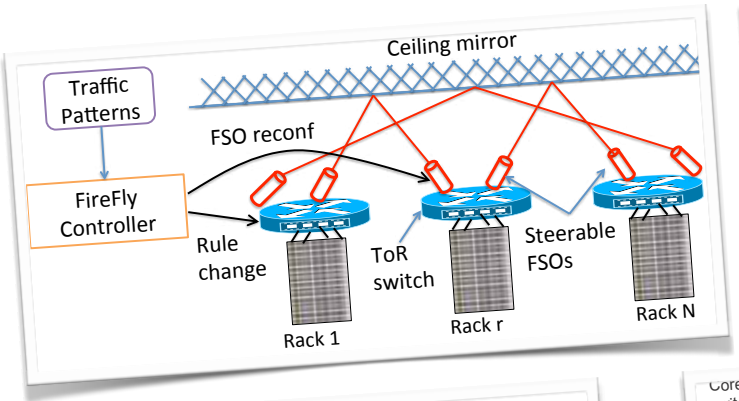
Skew in network traffic

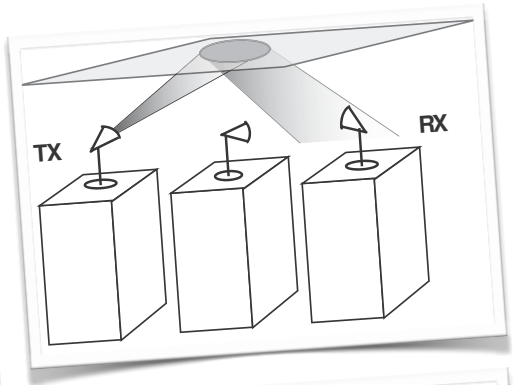
How does this observation help?

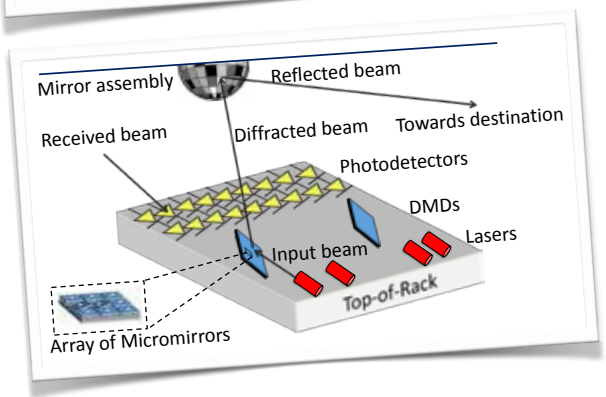
"Unfortunately, given current data center network architectures, the only way to provision required bandwidth bandwidth between dynamically changing sets of nodes is to build a non-blocking switch fabric at the scale of an entire data center"

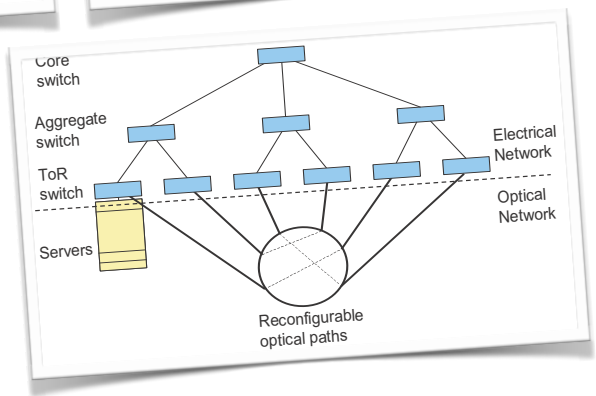
Helios: A Hybrid Electrical/Optical Switch Architecture for Modular Data Centers.
Farrington et al. ACM SIGCOMM 2010.

Set up network connections on the fly?









- OFC '09 Glick et al.
- SIGCOMM '10 Wang et al.
- SIGCOMM '10 Farrington et al.
- SIGCOMM '11 Halperin et al.
- NSDI '12 Chen et al.
- SIGCOMM '12 Zhou et al.
- SIGCOMM '13 Porter et al.
- SIGCOMM '14 Liu et al.
- SIGCOMM '14 Hamedazimi et al.
- SIGCOMM '16 Ghobadi et al.
- NSDI '17 Chen et al.

Can statically wired networks compete with this?

DC architecture @ SIGCOMM 2017

RotorNet: A Scalable, Low-complexity, Optical Datacenter Network

William M. Mellette, Rob McGuinness, Arjun Roy, Alex Forencich,
George Papen, Alex C. Snoeren, and George Porter
University of California, San Diego

Beyond fat-trees without antennae, mirrors, and disco-balls

Simon Kassing
ETH Zürich
simon.kassing@inf.ethz.ch

Asaf Valadarsky
Hebrew University of Jerusalem
asaf.valadarsky@mail.huji.ac.il

Gal Shahaf
Hebrew University of Jerusalem
gal.shahaf@mail.huji.ac.il

Michael Schapira
Hebrew University of Jerusalem
schapiram@huji.ac.il

Ankit Singla
ETH Zürich
ankit.singla@inf.ethz.ch

A Tale of Two Topologies: Exploring Convertible Data Center Network Architectures with Flat-tree

Yiting Xia, Xiaoye Steven Sun, Simbarashe Dzinamarira,
Dingming Wu, Xin Sunny Huang, T. S. Eugene Ng
Rice University

Rankings

ETH Zurich regularly features in international r
universities in the world and the leading univer

Ranking	2016
THE – World University Ranking, Times Higher Education	9th
QS – World University Rankings, Quacquarelli Symonds Ltd	8th
ARWU – Academic Ranking of the World Universities, Shanghai Jiao Tong University	19th